International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Interoperable Intelligent Environmental Decision Support Systems: a Framework Proposal

Miquel Sànchez-Marrè

Knowledge Engineeering & Machine Learning Group (KEMLG), Computer Science Dept.,
Universitat Politècnica de Catalunya-BarcelonaTech,
Jordi Girona 1-3, 08034 Barcelona, Catalonia
(miquel@lsi.upc.edu)

Abstract: In this paper, an approach for the development of Interoperable Intelligent Environmental Decision Support Systems (IEDSS) is proposed. The framework is based upon the cognitive-oriented approach for the development of IEDSS proposed in (Sànchez-Marrè et al., 2008), where three kind of tasks must be built: analysis tasks, synthesis tasks and prognosis tasks. Now, a fourth level will be proposed: the model construction layer, which is normally an off-line task. At each level, interoperability should be possible and inter-level interoperability must be also achieved. This interoperability is proposed to be obtained using data interchange protocols like Predictive Model Markup Language (PMML), which is a model interchange protocol based on XML language, using an ontology of data and AI models to characterize data types and AI models and to set-up a common terminology, and using workflows of the whole interoperation scheme. In the future, a Multi-Agent System will be used to implement the software components. An example of use of the proposed methodology applied to the supervision of a Wastewater Treatment Plant is provided. This Interoperable IEDSS framework will be the first step to an actual interoperability of AI models which will make IEDSS more reliable and accurate to solve complex environmental problems.

Keywords: Model Interoperability, Intelligent Environmental Decision Support Systems.

1. INTRODUCTION

Environmental systems/problems are real-world complex processes very difficult to manage and supervise. Most of them are dynamic processes with many decisions involved within them. For instance, the management of a basin river in a country taking into account all the environmental features: water level at different points, water quality of the river and its tributary rivers, wastewater treatment plants operation along the basin river, water use along the basin river, several stakeholders involved, politic and scientific strategies, etc. Another example could be air pollution control and management systems in charge of supervising the air quality of a concrete zone (region, city, etc.) taking into account all the features: car traffic, meteorological condition, chemical variables, physical conditions of the region, wind strength, etc. Environmental Decision Support Systems (EDSS) field has been trying to use some models of the real world being analysed to try to get an insight of the behaviour and evolution of the real system.

A model is a description of a system, usually a simplified description less complex than the actual system, designed to help an observer to understand how it works and to predict its behaviour. Typically, models could be divided into *mechanistic models* and *empirical models*. *Mechanistic models* are based on an understanding of the behaviour of a system's components, analysing the system from its first-principles. Usually these mechanistic models are expressed a as set of mathematical formulas and equations (differential equations, etc.). Historically, the first EDSS were using only these kind of models. Notwithstanding, taking into account that usually huge amount of data gathered from the system's being managed were available, some new empirical models were started to be used. *Empirical models* are based on direct observation, measurement and extensive data records. The first empirical models used were mathematical and statistical methods like Multiple Linear Regression (MLR) models, Principal Component Analysis (PCA) models, Discriminant Analysis (DA) models, Logistic Regression (LR) models, Statistical Clustering models, etc. The success of several inductive machine learning techniques, within the Artificial Intelligence area, lead to the use of another kind of empirical models: the intelligent data analysis models. Some instances are the Association Rules (AR) model, Classification Rules (CR) models, Decision Tree (DT) models, Artificial Neural Networks (ANN) models, Case-Based Reasoning (CBR) models, Fuzzy Logic (FL) models, Evolutionary Computation (EC) models, Bayesian Networks (BN) models, Conceptual Clustering (CC) models, etc. Since the 80s, both the former mathematical/statistical empirical models and the later machine learning empirical models have been named as *data mining methods*, because the models constructed are the result of a mining process among the data.

With the use of data mining models coming from the Artificial Intelligence field, the EDSS have been evolved to the Intelligent Environmental Decision Support Systems (IEDSS) (Sànchez-Marrè et al., 2006). IEDSS are built using some artificial intelligence model or integrating several artificial intelligence models to be more powerful, jointly with geographical information system components, mathematical or statistical models, and environmental/health ontologies, and some minor economic components. IEDSS integrate the expert knowledge/data stored by human experts through years of experience in the process operation and management. In addition, some knowledge/data can be mined through the intelligent analysis of large databases coming from historical operation of the environmental process. Thus, knowledge/data mining (model production), as well as reasoning over the produced models, and interoperation among the several models is a key step to build reliable IEDSS.

Single AI models provide a solid basis for construction of reliable and real applications, but the interoperability of AI/Numerical models is one of the main open challenges in this field.

2. RELATED WORK AND BACKGROUND

The Institute of Electrical and Electronics Engineers (IEEE) defined interoperability as “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” (IEEE, 1990). Additionally, when the components share a common understanding of the information model behind the data being interchanged, semantic interoperability is achieved (Manguinhas, 2010).

Semantic Integration and/or semantic interoperation has been the focus of some research works coming from environmental modelling field. Some pioneer work was done in semantic integration of Environmental models for application to global information systems and decision making, specially related to GIS components and models (Mackay, 1999; Wesseling *et al*., 1996). In addition, some work was done in model and data integration and re-use in EDSS (Rizzoli et al., 1998) and an overview of model integration was analysed in (Argent, 2004). An interesting work was presented in (Sottara et al., 2012) using the Drools Rule-based integration platform using a unified data model and execution environment.

In the information systems field several recent works were done in semantic integration of business components (Elasri and Sekkaki, 2013; Kzaz et al., 2010). Others interesting works focused on the semantic interoperability through service-oriented architectures (Vetere and Lenzerini, 2005).

In medical domains, also it has been the aim of several works using service-oriented architectures (Komatsoulis et al., 2008) or aiming at data interoperability on any Health Level Seven’s (HL7, http://www.HL7.org) standard (Dolin and Alschuler 2011).

Although there are some architecture proposals in the literature to combine some of these models, there is not a common framework to be taken into account as first guideline to deploy Interoperable IEDSS providing an easy way to integrate and (re)use several AI models or statistical/numerical models in a whole IEDSS. Until now, most of the interoperability of the models is achieved by a manual *ad hoc* model interaction. The aim of this current research work is to provide a useful and *systematic approach* to interoperate different models at different steps in the IEDSS solving process.

3. MODEL INTEROPERABILITY

In an IEDSS, regarding the models used there are two different steps: the *model production*, and the *model use*. Models are basically produced in a data mining process or directly from an environmental expert who provides the knowledge body of the model (inference rules, probability values, etc.). After

that step, the models can be used for *visualization purposes* (visualization a decision tree, visualization of a Bayesian network, visualization of a rule set, etc.) in order to validate and/or interpret the model by the experts, or for *execution purposes* (diagnosis of the current state of the environmental system, prediction of values for numerical variables, class discrimination of new instances, simulation of what-if scenarios, etc.).
This means that in a general IEDSS system, at least three different kind of processes must exist:

- Model producers, like for instance a decision tree mining algorithm which produces a decision tree model from raw data
- Model visualizers, like a tree visualizer algorithm which produces a tree shape image of the nodes and the leaves of the tree, from a decision tree model
- Model executors, like a decision tree interpreter which discriminates the class to which belongs a new instance or example, from a decision tree model

In addition, a model executor could also generate a new model, thus adopting the role of model producer as well. This process is depicted in figure 1.

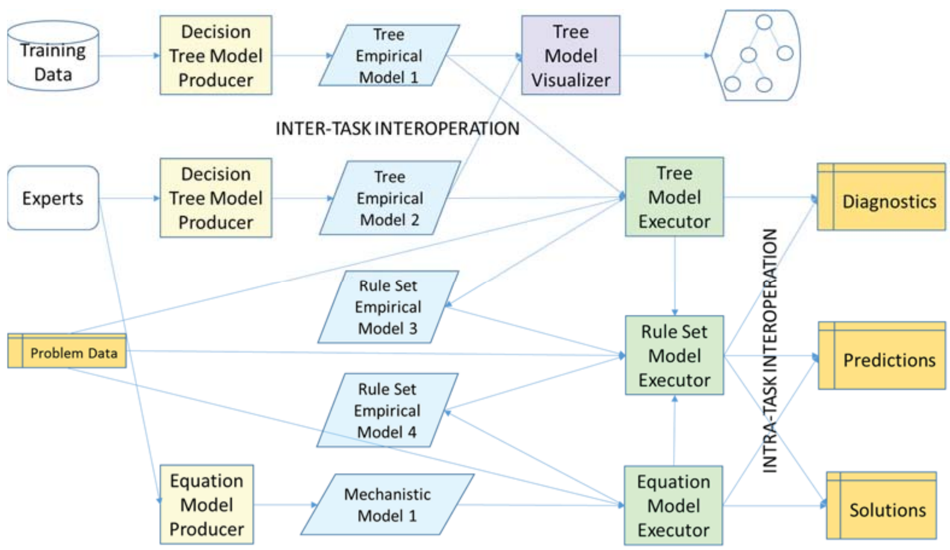

Figure 1. Inter-task interoperation and intra-task interoperation types

All three model tasks (production, visualization, execution) must be able to share the models and have a clear understanding of them. To that end, what we are proposing is to use a model interchange format, which will be based on the Predictive Modelling Markup Language (PMML), developed by the Data Mining Group (DMG) to achieve the inter-task interoperation between these three model tasks. Also, interoperability must be achieved inside the same task category. Model executors can generate new models, and mainly can activate other model executors. Thus, intra-task interoperability must be accomplished too. With the same model interchange format plus the possibility to activate/call other model executors this will be accomplished.

This means that one model could be produced by one data mining tool, and could be visualized by another visualizing tool, and can be executed by another model interpreter tool, given that all are compliant with the PMML model interchange format. This really is a very powerful ability to use the most convenient or desired tool for the environmental scientist. Notwithstanding, the approach that will be presented in section 5 proposes the same tool for all the different task models.

4. TOOLS AND TECHNIQUES FOR ACHIEVING INTEROPERABILITY

To achieve interoperation at both inter-task model and intra-task model, the following tools, standards and techniques are proposed to be used.

One of the most effective ways to interchange information between several software components, and to share the semantics associated to the information is XML (eXtensible Markup Language). XML is a meta-language intended to supplement HTML's presentation features with the ability to describe the nature of the information being presented (Erl, 2004). XML adds a layer of intelligence to information being interchanged, providing meta-information, which is encoded as self-descriptive labels for each piece of text that go wherever the document goes. XML is implemented as a set of elements, which can be customized to represent data in unique contexts. A set of related XML elements can be classified as a vocabulary. An instance of a vocabulary is an XML document. Vocabularies can be defined formally using a schema definition language like Document Type Definition (DTD) or XML Schema Definition Language (XSD).
The Data Mining Group (DMG, 2014) is an independent, vendor led consortium that develops data mining standards, such as the Predictive Model Markup Language (PMML). PMML is a standard for statistical and data mining models and supported by over 20 vendors and organizations. PMML uses XML to represent mining models. The structure of the models is described by an XML Schema. One or more mining models can be contained in a PMML document. A PMML document is an XML document with a root element of type PMML. A PMML document can contain more than one model, and most common data mining models are supported (AssociationModel, RegressionModel, TreeModel, RuleSetModel, NeuralNetwork, ClusteringModel, etc.).The general structure of a PMML document is:

```
<?xml version="1.0"?>
<PMML version="4.2"
  xmlns="http://www.dmg.org/PMML-4_2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">

  <Header copyright="Example.com"/>
  <DataDictionary> ... </DataDictionary>
   <RuleSet model … >
    ... a model ...
   </RuleSetModel>
    . . .
</PMML>
```

Visual workflow building and execution will be a very useful tool for specifying the workflow involving the raw data, the models produced, the model executors, and auxiliary processes. Workflows are graphical notations, which were first introduced to model and describe Business Processes. They let the design and specification of a workflow involving several elements (data, interchange PMML models, model producers, model executors, solution combiners, current problem specification, etc.). In our approach, the use of jBPM will be assessed. It is an open source business process engine supporting the Business Process Model and Notation (BPMN 2.0) standard, and provide a graphical workflow editor. The idea is that the solving process of the IEDSS through its different tasks and layers (data mining, diagnosis tasks, solution-generation tasks and predictive tasks) will be described through the workflow, and the workflow will be directly executed or perhaps will occasionally generate a corresponding software code, which can be minimally modified.

The use of an ontology of the several AI models supported by the whole framework will be convenient and probably necessary to assist in the automatic execution of the workflows. An ontology is a formal description and characterization of the knowledge about some domain. It defines the important concepts and entities, their relevant features and their relationships. The ontology can be describing not only the AI models but also the mathematical/statistical models used in the IEDSS developed.

In the prognosis layer, where several alternative solutions must be evaluated to give the corresponding feedback to the IEDSS user, our approach proposes that numerical simulation models, qualitative simulation models and an agent-based simulation environment tool to model and evaluate the complex scenarios that must be assessed in the IEDSS (Rendón-Sallard et al., 2006). A proposed multi-agent platform tool will be JADEX (Braubach et al., 2003). Jadex is an addon to the widely used JADE agent platform. The add-on follows the BDI architecture, which is a well-known model for representing mentalistic concepts in the system design and implementation.

5. THE INTEROPERABLE IEDSS FRAMEWORK

In (Sànchez-Marrè et al., 2008) a cognitive-oriented approach for IEDSS development was proposed. This architecture was a three-layer one distinguishing the analytical tasks, synthesis tasks and prognosis tasks performed in the problem solving of an IEDSS. Having in mind that the separation between model production and model execution is an important feature to achieve interoperability it is proposed to design a four-layer cognitive-oriented approach, were three sub-layers corresponding to three tasks mentioned before:

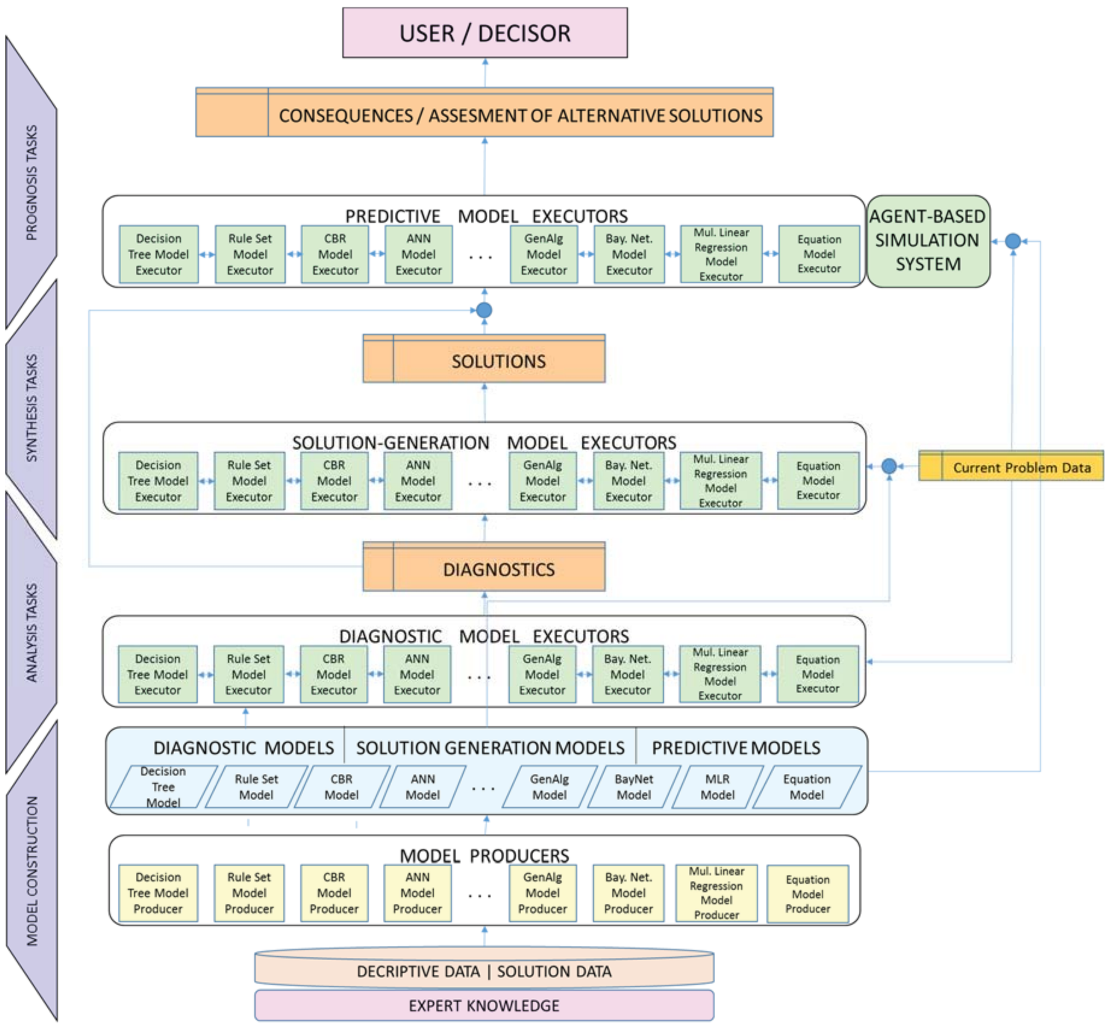

Figure 2. Four-layer IEDSS development architecture

- Model Construction layer: At this stage the data gathering processes as well as the knowledge discovery process by means of some data mining techniques or acquiring the expert knowledge are undertaken to get both diagnostic models, solution-generation models and predictive models.
- Model Execution layer: This layer include the analysis tasks, synthesis tasks and prognosis tasks. This layer is compose by three-sub-layers:
  - Analysis task layer: In this layer is where most of interpretative processes are run. The Diagnosis models are executed here to provide the IEDSS with hard analytical power to get an insight of the environmental system being supervised in real-time or managed in an off-line basis.
  - Synthesis task layer: This layer wraps all the work necessary to synthesizes possible alternative solutions for the different diagnostics found in the previous step. This

synthetisation can be done through several solution-generation models produced on the model construction layer. Those models could be Mathematical/Statistical models or AI models, and even mechanistic models. Of course, the integration and interoperability of different nature models would enhance the problem solving ability of the IEDSS.

- Prognosis task layer: At this upper layer relays the inherent ability of IEDSS to decision support tasks. At this level, the predictive models constructed at the model construction layer, which can be numerical (mostly simulations) or rather qualitative (qualitative reasoning or qualitative simulations), are used to estimate the consequences of several actions proposed in the previous layer by the solution-generation models. These "what if" models let the final user/decisor to make a decision based on the evaluation of several possible alternatives. At this stage, the temporal and spatial features could be very important for a good environmental modelling practise.

Figure 2 depicts the four-layer IEDSS development architecture.This framework will be deployed and implemented over GESCONDA tool (Sànchez-Marrè et al., 2010). This software is an Intelligent Decision Support System tool in development in the KEMLG group at Universitat Politècnica de Catalunya-BarcelonaTech. GESCONDA is a four-layer architecture enabling Data filtering and preprocessing, Recommendation and Meta-Knowledge Management, Data Mining & Knowledge Discovery, and Knowledge Management and Reasoning. Until recent years, the tool was just a data mining tool, but with the addition of its fourth level, some reasoning abilities were provided. The tool has a Case-Based Reasoning engine and a Rule-Based Reasoning engine, which are able to execute rule set models and Case-Based models. Currently an evolutionary computing engine is being implemented. Therefore, the tool is able to both produce models and to execute models. What is now being started to be done is to use the PMML standard and make the GESCONDA tool compliant with it.

6. THE USE OF THE FRAMEWORK IN A CASE STUDY

In this section, a possible use of the interoperable IEDSS framework in the supervision and management of a wastewater treatment plant is described. Biological Wastewater Treatment Plant (WWTP) is a complex process that involves chemical and biological reactions, kinetics, catalysis, transport phenomena, separations, and so on. The quality of the effluent water must be always maintained in a good condition to minimise any environmental impact. Nevertheless, some features such as inflow changes, both in quantity and in quality, and the population variation of the microorganisms over time, both in quantity and in the relative number of species, makes the process very complex. In addition, the process generates a huge amount of data from different sources (sensors, laboratory analyses and operator's observations), but these data are often uncertain, subjective or vague. In the wastewater treatment plant (WWTP) operation, several problems frequently appear such as solids separation problems, biological foam or bulking episodes in the bioreactors, or overloading derived from storms and heavy rains. Usually there are three control variables like the Dissolved Oxygen (DO in mg/l) in the bioreactor, the Recirculation Flow (RF in m^3/day) and the Waste Flow (WF in m^3/day). Let us suppose an actual case were data from one WWTP are available. Each data entry is a mean value of several variables, and suppose that we wanted first to get a set of clusters or classes of the different data entries to find the different operation situations. After that, we wanted to obtain a rule-set model for diagnosing the operation situation (class) to which belongs a new data entry (rule-set model). In addition, we wanted to obtain a decision-tree model for getting the same diagnosis, to compare and validate the diagnosis. An experience-based model (case-base model) both for diagnosis and solution generation is also required, and when using the Case-Based model to generate a possible solution for a concrete problem data, different solution-generation models are desired to be executed from the Case-based Model. For instance, the DO value is required to be computed using a Multiple Linear Regression (MLR) model and both the RF and WF values will be computed through the standard procedures of the Case-Based model. Finally, both the Rule-based solutions and the Case-based solutions will be merged to get a final possible solution to the current problem.

This scenario can be solved in a manual way with the GESCONDA tool, but in the next research efforts with the adoption/addition of a graphical workflow tool, the whole decision support process could be automatized. In figure 3 there is the chart of the interactions and interoperations between the software components. The automatization will be provided by the direct execution of a problem-solving workflow.

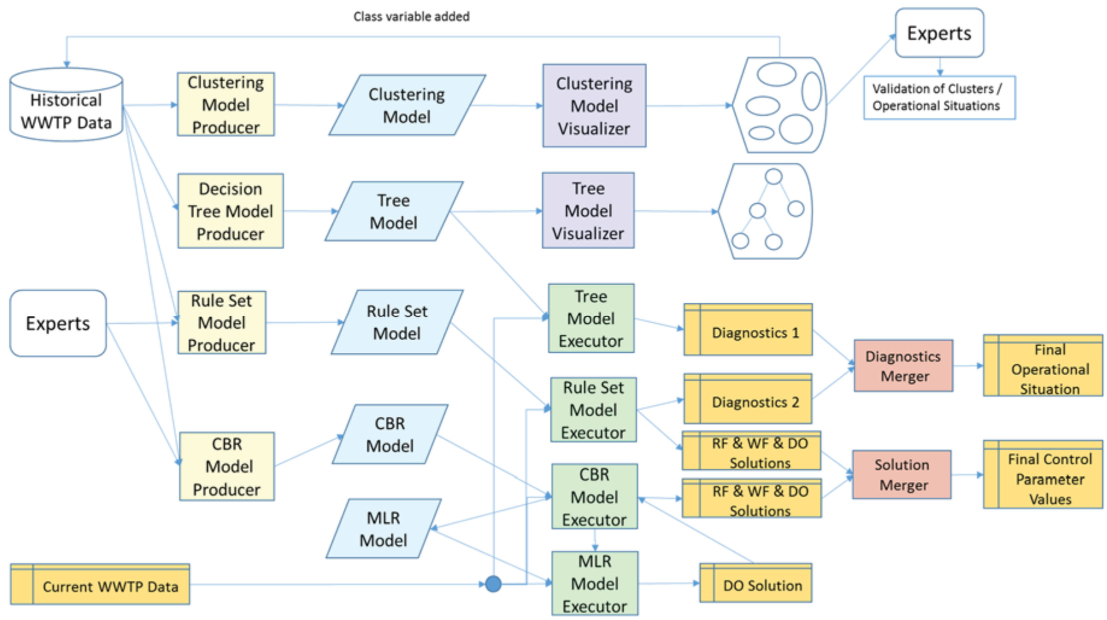

Figure 3. Chart of the workflow for the concrete scenario

7. SUMMARY AND DISCUSSION

In this paper, an approach to a possible architecture for developing Interoperable IEDSS has been presented. One of the main open challenges in current research in Intelligent Environmental Decision Support Systems is the interoperation and integration of several AI and Mathematical models within the same system. Usually, this interoperation is done manually, and in an *ad hoc* model interaction. In addition, there are not many available IEDSS tools, because most of them are just data mining tools, which can produce some models, but not the execution of the models. The framework architecture propose to have different software components: *model producers*, *model visualizers* and *model executors*. These software components will share the models, which will be expressed in a common interchange format, based in XML language. It is proposed to use the standard predictive model markup language (PMML) or an extended version of it to achieve the interoperation. Interoperation should be achieved both in *inter-task interoperation* among the different model software components (producers, visualizers, executors) and in *intra-task interoperation* among the same type of model software components. This approach provides a great flexibility to use different tools for the different tasks involved with the models. Thus, for example one model could be produced with one compliant interoperable tool like KNIME tool (KNIME, 2014) could be visualized with another compliant interoperable tool, and could be executed with a third different compliant interoperable tool. The proposed interoperable IEDSS framework is a four-layer cognitive-oriented approach: model construction layer, analysis task layer, synthesis task layer and prognosis task layer. The methodology proposed has been illustrated with a real scenario to manage a WWTP. The preliminary evaluation of the approach with some environmental experts has been positive.
In the future, all the different layers of the architecture, the interchange protocol formats, and a visual workflow to automatize the interoperation capabilities will be implemented and integrated in the GESCONDA IEDSS tool. Additionally, the use of a Multi-agent approach could be used not only for the prognosis level, but also for all the other task layers. First, the visual workflow approach to interoperability will be explored. In a second step, the multi-agent approach will be used.

REFERENCES

Argent, R.M. (2004). An Overview of Model Integration for Environmental Applications-components, frameworks and semantics. *Environmental Modelling & Software* 19:219-234, 2004.

BMPN (2011). The Business Process Model and Notation BMPN v2 (http://www.bpmn.org/, http://www.omg.org/spec/BPMN/2.0/) was released in January 2011.

Braubach, L, Lamersdorf, W. and Pokahr, A. (2003). Jadex: Implementing a BDI-Infrastructure for JADE Agents. EXP - In Search of Innovation (Special Issue on JADE), Vol 3, Nr. 3, Telecom Italia Lab, Turin, Italy, September 2003, pp. 76-85.

DMG (2014). The Data Mining Group (http://www.dmg.org) leads the development of the Predictive Model Markup Language (PMML). Current version PMML 4.2. February 2014.

Dolin,R.H. and Alschuler,L. (2011). Approaching semantic interoperability in Health Level Seven. *Journal of American Medical Informatics Association*, Vol. 18:99-103.

Elasri,H. and Sekkaki, A. (2013). Semantic Integration process of Business Components to Support Information System Designers. *Int. Journal of Web & Semantic Technologies* 4(1):51-65.

Erl, T. (2004). *Service-Oriented Architecture. A Field Guide to Integrating XML and Web Services*. Prentice-Hall, 2004.

Institute of Electrical and Electronics Engineers (1990). IEEE Standard Computer Dictionary: a Compilation of IEEE Standard Computer Glossaries. New York.

jBPM (2014). The jBoss Business Process Management engine jBPM v6.0 (http://www.jboss.org/jbpm) was released in January 2014.

KNIME (2014). Konstanz Information Miner (http://www.knime.org/). Current version KNIME 2.9.2, March 2014.

Komatsoulis G.A., Warzel, D.B., Hartel, F.W., Shanbhag, K., Chilukuri, Ram, Fragoso G., de Coronado S., Reeves, D.M., Hadfield, J.B., Ludet, C. and Covitz, P.A. (2008). caCORE version 3: Implementation of a Model Driven, Service-Oriented Architecture for Semantic Interoperability. *Journal of Biomedical Informatics* 41(1):1-29.

Kzaz, L., Elasri, H., and Sekkaki, A. (2010). A Model for Semantic Integration of Business Components. *Int. Journal of Computer Science & Information Technology* 2(1):1-12.

Manguinhas, H. (2010). Achieving Semantic Interoperability using Model descriptions. *Bulletin of IEEE Technical Committee on Digital Libraries*, Vol. 6, No. 2, Fall 2010.

Mackay, D. S. (1999). Semantic Integration of Environmental Models for Application to Global Information Systems and Decision-Making. *ACM SIGMOD Record* 28(1):13-19, March 1999.

Ouksel, A.M. and Sheth, A. (1999). Semantic Interoperability in Global Information Systems: a Brief Introduction to the Research Area and the Special Section. *ACM SIGMOD Record* 28(1):5-12, March 1999.

Rizzoli, A. E., Davis, J.R. and Abel, D.J. (1998). Model and Data Integration and re-use in Environmental Decision Support Systems. *Decision Support Systems* 24:127-144, 1998.

Rendón-Sallard, T., Sànchez-Marrè, M., Aulinas, M and Comas, J. Designing a Multi-Agent System to Simulate Scenarios for Decision Making in River Basin Systems. Ninth International Conference of the Catalan Association of Artificial Intelligence (CCIA'2006). *Frontiers in Artificial Intelligence and Applications Series, Vol. 146*. Artificial Intelligence Research and Development (Ed. M. Polit, T. Talbet, B. López y J. Meléndez), pp. 291-298. ISBN 1-58603-663-7. 26-27 de October 2006.

Sànchez-Marrè, M., Gibert, K. and Sevilla, B. (2010). Evolving GESCONDA to an Intelligent Decision Support Tool. 5th International Congress on Environmental Modelling and Software (iEMSs'2010). iEMSs' 2010 Proceedings, Vol. 3, pp. 2015-2024. ISBN 978-88-903574-1-1.

Sànchez-Marrè, M., J. Comas, I. Rodríguez-Roda, M. Poch and U. Cortés. (2008). Towards A Framework for the Development of Intelligent Environmental Decision Support Systems. 4th International Congress on Environmental Modelling and Software (iEMSs'2008). iEMSs' 2008 Proceedings, pp. 398-406.

Sànchez-Marrè, M., Gibert, K., Sojda R., Steyer J.P., Struss, P. and Rodríguez-Roda, I. (2006). Uncertainty Management, Spatial and Temporal Reasoning and Validation of Intelligent Environmental Decision Support Systems. 3rd International Congress on Environmental Modelling and Software (iEMSs'2006). iEMSs' 2006 Proceedings, pp. 1352-1377. ISBN 1-4243-0852-6.

Sottara, D., Bragaglia, S., Mello, P., Pulcini, D., Luccarini, L and Giunchi, D. (2012). Ontologies, Rules, Workflow and Predictive Models: Knowledge Assets for an EDSS. 6th International Congress on Environmental Modelling and Software (iEMSs'2012). iEMSs' 2012 Proceedings, pp. 204-211. ISBN: 978-88-9035-742-8.

Vetere, G. and Lenzerini, M. (2005). Models for Semantic Interoperability in Service-oriented architectures. *IBM Systems Journal* 44(4):887-903, 2005.

Wesseling, C.G., D. Karssenberg, P.A. Burrough, and W.P.A. Van Deursen. (1996). Integrating dynamic environmental models in GIS: the development of a dynamic modelling language. *Transactions in GIS*, 1(1), 40-48.